SECTION 4

PRELIMINARY STATISTICAL ANALYSIS OF DISPARITIES IN DEFAULT RATES

4.1. Statistical Tests of Geographical Differences

Although the nonstatistical work in Section 3 cannot determine whether there is "too much" geographical concentration of defaults, it does provide a useful beginning: a casual analysis of raw default rates that seems to indicate substantial concentrations in particular areas. It also took a reasonable second step, showing that there are numerous systematic factors that vary across tracts and which may be able to account for the observed intertract disparities in default rates, if in fact a statistical analysis determines that randomness alone is unlikely to be the explanation.

We now take up the question of whether the geographic concentration of defaults is likely to be attributable to chance, or whether the principal explanation lies in systematic factors, such as poor lending practices in certain neighborhoods or the factors examined in the last section. In a sense, then, we now take a step backwards. That is, despite having seen evidence that measurable default-related factors vary across tracts and might possibly account for variation in default rates across tracts, we now focus on the possibility that chance alone is responsible for the observed variation in default activity. The reason for taking this step is that, were one to attempt to isolate potential problem areas in practice, the first step might well be the one taken in this section: a simple statistical analysis of raw default rates. Hence, this section serves to illustrate these methods, including some of the problems inherent in applying these methods in practice.

It bears emphasis that here we depart in essential ways from the methods employed in the NTIC study, in which ratios of raw tract-level default rates to MSA rates are taken as proof of underlying problems. In contrast, the current study accounts for randomness in outcomes in deciding whether defaults are indeed too concentrated. An approach that acknowledges the existence of randomness is surely reasonable, for default rates are the product of a myriad random effects that could lead to sharp divergence in default probabilities across tracts.

A primary advantage of the statistical approach is that it recognizes the influence of randomness in a precise way. That is, it is not simply that the nonstatistical approach will necessarily identify too many tracts (or lenders) as high-default entities, and that if one simply relaxes the statistical standards for identifying tracts as high-default (by raising the allowable probability that chance alone is responsible) one will end up with the same set of tracts as would be obtained with the nonstatistical approach embraced by the NTIC study. Instead, the nonstatistical NTIC approach may single out some low-volume tracts as high-default where chance alone is a likely explanation, but it may ignore other high-volume tracts where chance alone is very unlikely to cause a default rate as high as that observed (but not high enough to reach the NTIC cutoff). That is, the NTIC approach allows for randomness in an inconsistent manner, and the result of using such methods to identify high-default tracts and prescribe solutions is an incorrect focus that is less likely to discover tracts where there truly are systematic causes for defaults, such as poor underwriting or inadequate servicing of delinquent borrowers.

We emphasize that the focus of the statistical approach in essentially ignoring tracts for which randomness appears to be a plausible explanation for default activity does not in any way minimize the importance of defaults caused by chance factors alone. Foreclosure-induced vacancies may be a problem for neighborhoods regardless of the cause of foreclosure. The reason for focusing on tracts where there appear to be systematic factors, rather than solely random factors, causing default activity is that systematic factors are more likely to be permanent and more likely to be amenable to identification and remedial action.

At the same time, a blind application of statistical methods has its own drawbacks in the current context. In particular, looking only at the probability that chance alone is responsible for the level of defaults recognizes statistical significance, but says nothing about practical significance. In MSAs that are characterized by low average default rates, like Denver (see Table 1 in Appendix A), tracts may be properly singled out as high-default because their default rate is very unlikely to be produced by chance alone, yet the tract default rate is low enough to be of little practical importance. This problem does not argue in favor of the nonstatistical NTIC approach; it argues in favor of recognizing the gains from reducing the tract-level default rates as an additional criterion. In this study we crudely recognize the importance of possible gains from default rate reductions by paying most of our attention to entities (tracts or loans) with more than 30 loans. A superior alternative in practice may be to isolate problem tracts based on both statistical significance and the expected gains from reducing tract default rates.[20]

Before presenting the statistical work, we must also return to an issue raised in the introduction to this paper. As noted there, we shall use statistical procedures to ask *how unlikely* it is that we would find a particular outcome if in fact chance alone were at work. Although the statistical analysis will provide us with an answer to the latter question, it cannot answer for us how unlikely is *too* unlikely for us to accept the notion that chance alone is at work. If, for example, we find that there is only a 5 percent probability that we would obtain an outcome as extreme as that observed if chance alone were at work, we must still decide if 5 percent is so low a probability that we would reject chance as the explanation. Although there is no easy answer to this question, we will generally adopt the (essentially arbitrary) convention often used in empirical work: we shall reject chance alone as an explanation if we find that there is a 5 percent probability, or less, that an outcome as extreme as what we observe could have been produced by chance alone. Stated in traditional statistical terms, we shall agree to reject the hypothesis at issue (*e.g.*, that tract-level or lender-level default probabilities equal MSA default rates) at a significance level of 5 percent.[21]

4.1.1. Chi-Square Tests of Geographical Differences

To begin, we perform a standard chi-square test for whether the incidence of defaults (under any the three definitions) is independent of tract, given the number of loans in each tract and the overall number of defaults and nondefaults in the MSA as a whole. As we shall see, this test has

[20] We do not mean to deny the importance of default at the local level. Indeed, one of the gains from a reduction in defaults is presumably the localized benefit to the neighborhood.

[21] The hypothesis tests in this study are generally one-tailed tests. The choice of a smaller significance level could surely be justified as well.

limited applicability in the current study because of sample size problems. We present the results nonetheless, for such tests are commonly applied in situations like these, and they may be used to advantage in other investigations of default data. The approximations employed in this test are better met when there are adequate expected cell sizes, and thus we follow one convention by including only those tracts for which the expected number of defaults (calculated as the number of loans in the tract times the MSA default rate) is at least 5 (see Fleiss [1981]).[22] In these data, this standard is generally far more demanding than a requirement that there be at least 30 loans in a tract. Indeed, the effect of this restriction is to make many tracts ineligible for inclusion in the test by themselves, especially when defaults are defined to include only claims at two years --- a very rare event.

The results of these tests under the three default definitions for both origination years pooled[23] are given in the three panels of Table 12 (in Appendix A). The first two columns of numbers in each panel show the value of the chi-square statistic and the number of degrees of freedom, the latter of which equals the number of tracts included in the analysis minus one. The final column in each panel shows the probability that the intertract pattern in defaults could have arisen from chance alone, *i.e.,* the probability of observing this extreme an outcome if there really is no association between default behavior and tract identity.

As indicated above, the number of tracts included in this analysis (the number of degrees of freedom plus one) is generally far smaller than the total number of tracts in the MSA (reported in Table 2 above). This problem renders the analysis questionable at best, particularly for the claims definition utilized in Panel A. In those few MSAs where the number of tracts included in the analysis is at all large, we generally see probabilities well below 0.01 (1 percent) that a pattern so extreme could have arisen from chance alone. Putting aside the meaning of a test in which so many tracts are eliminated from consideration, we would generally be led to reject the hypothesis that defaults are distributed independently across tracts.[24]

4.1.2. Exact Probability Calculations

The last set of tests offers very weak evidence that differences in default propensities across tracts are unlikely to have arisen from chance alone. As noted, substantial numbers of tracts are excluded from the tests above because they have so few loans that they would otherwise render the statistical approximations too inaccurate for comfort. In addition, even if the latter difficulty did

---

[22] When a tract has too few loans for inclusion in the calculation as an individual tract, we put it in a pool that is ultimately used as a separate tract in the chi-square calculation. One alternative to this procedure is to omit such low volume tracts from the analysis entirely. It is unclear which alternative is superior.

[23] Analyzing each origination year separately, which would be desirable in other respects, exacerbates the problem of small sample sizes.

[24] A Fisher exact test would be preferable. The latter test gives the exact probability of observing an outcome at least as extreme as that observed and can contend with cells of any size. Such a test is extremely calculation-intensive, particularly on tables of the size used here, and for this reason, we have not implemented these tests.

not arise, these tests alone do not tell us in which particular tracts the problems may lie. That is, these tests do not isolate the tracts in which defaults differ significantly from what would be expected on the basis of chance alone.

To get around these problems, we present a second series of tests. In contrast to the last tests, these are conducted on a tract-by-tract basis. We calculate the exact probability of obtaining the number of defaults observed in a tract, or more, if in fact defaults occur with the same probability within a tract as at the MSA level.[25] The problem with these tests, however, is that there is no recognition of the interdependence across tracts; that is, the tests as a group do not account for the fact that, with fixed margins at the MSA level, increases in defaults in one tract necessarily reduce the number that must be distributed across other tracts.

The outcomes of these "exact probability calculations" are given in the various panels of Table 13; each panel corresponds to an alternative definition of default and a different origination year combination. Panel F is included below and is the basis for some of our discussion; the remaining panels of Table 13 may be viewed in Appendix A. We present the total number of tracts in the first column, and in the remaining columns we show the fraction of tracts falling into each "Probability of Outcome" category. The latter probability is that of obtaining as many defaults as observed in a tract, or more, if in fact the tract-level default probability is the same as that at the MSA level.[26] These calculated probabilities are separated into three groups: 5 percent or less, 5 to 10 percent, and greater than 10 percent.[27] Hence, tracts falling in the first category (5 percent or less) are those for which the probability of obtaining as many or more defaults as that observed (assuming chance alone is responsible) is 5 percent or less; those in the second group are tracts for which the probability is more than 5 but less than 10 percent; and so on. In the first row of Panel F of Table 13, for example, we see that 6.93 percent of the 476 tracts in the Atlanta MSA have a probability of up to 5 percent, another 3.99 percent have a probability of more than 5 but no more than 10 percent, and the remaining 89.08 percent have a probability of more than 10 percent. Looking across the various panels of Table 13, we see that there are quite clearly numerous tracts in which the probability of obtaining as many or more defaults from chance alone is 5 percent or less, an observation that seems to reinforce the idea that the geographic distribution of defaults is unlikely to be generated by random forces, but instead there are real intertract differences in default probabilities.

There is, however, another interesting feature of Table 13 that deserves explicit mention. Even if there is a very low (less than 5 percent) probability that so many defaults would be generated within a tract if chance alone were at work, there is still *some* probability that randomness

---

[25] We use a binomial model, which entails an assumption of sampling with replacement. When calculating probabilities for an individual tract, the latter assumption is unlikely to be of great importance. That is, the alternative assumption of sampling without replacement, which is based on the hypergeometric distribution rather than the binomial, is likely to yield results that are very similar to those presented above.

[26] We may alternatively think of this as a significance level category for a test of the null hypothesis that the tract-level and MSA-level default probabilities are the same, against the alternative that the tract-level probability is higher.

[27] The analysis is performed on only those tracts with at least 2 loans in the origination years under consideration. Tracts with only one loan are guaranteed to fall in the third category when the MSA default rate exceeds 10 percent, and thus we exclude single-loan tracts from all calculations.

TABLE 13

EXACT PROBABILITY CALCULATIONS ASSUMING THAT TRACT DEFAULT PROBABILITY EQUALS MSA DEFAULT RATE FOR EACH TRACT

PERCENTAGE OF TRACTS BY PROBABILITY OF OUTCOME , BY MSA

PANEL F: 1992 AND 1994 ORIGINATIONS, UNCURED DELINQUENCIES AT TWO YEARS

MSA NAME	TOTAL TRACTS	PROBABILITY OF OUTCOME		
		5% OR LESS	5% TO 10%	>10%
ATLANTA, GA MSA	476	6.93%	3.99%	89.08%
BALTIMORE, MD PMSA	460	6.74	2.83	90.43
CHICAGO, IL PMSA	1083	5.91	2.95	91.14
DALLAS, TX PMSA	489	5.93	2.66	91.41
DENVER, CO PMSA	396	4.55	2.02	93.43
DETROIT, MI PMSA	941	4.46	1.06	94.47
FORT LAUDERDALE, FL PMSA	140	6.43	2.14	91.43
FORT WORTH-ARLINGTON, TX PMSA	295	5.08	3.73	91.19
HOUSTON, TX PMSA	499	2.61	2.20	95.19
LOS ANGELES-LONG BEACH, CA PMSA	729	3.84	2.61	93.55
MEMPHIS, TN-AR-MS MSA	187	9.63	2.67	87.70
MIAMI, FL PMSA	203	5.91	2.96	91.13
MINNEAPOLIS-ST PAUL, MN-WI MSA	612	5.88	1.63	92.48
NORFOLK-VIRGINIA BEACH-NEWPORT	311	7.07	2.57	90.35
ORLANDO, FL MSA	215	7.44	2.33	90.23
PHILADELPHIA, PA-NJ PMSA	726	3.17	2.75	94.08
PHOENIX-MESA, AZ MSA	475	4.84	3.16	92.00
RIVERSIDE-SAN BERNARDINO, CA PMSA	316	6.33	2.85	90.82
SACRAMENTO, CA PMSA	255	2.35	3.92	93.73
ST. LOUIS, MO-IL MSA	366	5.74	1.64	92.62
TAMPA-ST PETERSBURG-CLEARWATER	340	3.82	2.35	93.82
WASHINGTON, DC-MD-VA-WV, PMSA	755	4.90	2.78	92.32

alone is responsible for the observed disparity, *i.e,* that the tract-level default probability differs from the MSA rate only because of randomness. When conducting numerous tests of this kind, one expects to encounter some cases in which the unusual does in fact occur. A finding that some tract-level tests fail our standard is thus to be expected, and there is substantial risk in attributing much meaning to a finding that *some* tracts fail statistical tests. Indeed, the adoption of a 5-percent standard will, when applied independently to a large number of tracts, generate rejections that randomness is responsible in about 5 percent of tracts, even if it is true that only randomness is at work. Such is the fallible nature of statistical tests.

In Table 13, then, one might expect to find the second column containing about 5 percent of tracts and the third column another 5 percent even if randomness prevails. In Table 13 we find that the percentage of tracts in the second and third columns is typically less than five percent when defaults are measured as claims at two years, and thus these findings often appear consistent with the pattern anticipated if defaults occurred randomly within tracts at the same rate as in the MSAs as a whole. In other panels, especially those that pool the two origination years and measure defaults as uncured delinquencies (Panels F and I), we often find more than 5 percent of the tracts falling in the column pertaining to a significance level of 5 percent. The latter finding suggests that defaults occur too frequently in a disproportionate share of the tracts compared to what would be expected from chance alone.

It should also be noted that once again patterns for the claims only definition of default appear to differ somewhat from patterns that emerge for the other two default definitions. The differences are again consistent with the notion that claims at two years are disproportionately those defaults that are traceable to purely random factors.

Comparisons of the probability calculations for the two different origination years are also revealing. Table 14 (in Appendix A) presents these results. The first column of numbers in each row gives the number of tracts that appear in both origination years. The remaining columns give the percentage of these tracts for which exact probability calculations in Table 13 (for a particular origination year) yield a number of 5 percent or less, thus identifying the tract as a "high-default tract" in that year by our current standards. In the first row of Panel B, for example, we see that among the 384 Atlanta MSA tracts appearing in both origination years, 0.78 percent of the tracts are identified as high-default tracts in both years and 6.77 percent are identified as high-default tracts in only one of the two years. As might have been expected from the findings in Table 5 (in Appendix A), we see that there is a substantial change in the identity of tracts that test out as high-default tracts, which again reinforces the notion that at least some of the effects being picked up here are transitory.[28]

- **To conclude, this statistical work suggests that in many tracts default rates are too high to be plausibly explained by chance alone, though the fraction of such high-default tracts varies substantially from MSA to MSA. The identities of these tracts vary greatly with the year of origination, however, casting some doubt on the importance of the disparities in default rates.**

---

[28] If effects were purely transitory, in the sense of independent across years, we would expect to see that only 5 percent of the tracts singled out as high-default in one year would also appear as high-default in the other.

A more detailed statistical treatment presented in Section 5 will help reveal whether the apparent disparities hold up once we control explicitly for observable differences among the loans and borrowers within the tracts that are tentatively labeled as high-default.

4.2. Statistical Tests of Differences Across Lenders

4.2.1. Chi-Square Tests of Differences

We begin in an analogous way to that in the last section: by testing for independence of defaults across lenders within each MSA, holding fixed the total number of defaults in the MSA as well as the number of loans made by each lender. Table 15 (in Appendix A) shows the chi-square statistics[29] and the associated probability that these data could have been generated by chance alone if in fact there were no association between lender and default behavior. We see that, as was the case with tracts, very few lenders (relative to the number potentially available, as reported in Table 9) are included individually in the MSA-specific analyses, especially when defaults are defined to include only claims (Panel A). Once again, this fact casts serious doubt on the usefulness of such a series of tests. Where the analysis includes reasonably large numbers of lenders, we generally find very low probabilities (much lower than 0.01, or 1 percent) that the observed pattern could have arisen from chance alone. Hence, the evidence suggests --- again weakly --- that defaults are concentrated in lenders beyond what could reasonably be attributed to purely random forces.

4.2.2. Exact Probability Calculations

As in the last section, we now calculate the probability that each lender would have as many defaults as observed, or more, under the assumption that defaults for a lender occur randomly with the same probability as in the MSA as a whole. Table 16 reports the results of these calculations. Panel F is given below, and the remaining panels are in Appendix A. As was the case with an examination of census tracts, we see that some lenders fail the test and are thus identified as "high-default lenders" by the current standard.

As was the case with tracts, we see that in some panels of Table 16, especially those for which defaults are measured as claims at two years, there are generally no more than 5 percent of the lenders in the column that corresponds to a "probability of outcome" of 5 percent and no more than another 5 percent in the column that corresponds to a "probability of outcome" between 5 and 10 percent. As was the case with tracts, however, there are often more than 5 percent of the lenders in the column with a 5 percent "probability of outcome" when defaults are measured as uncured delinquencies, especially when both origination years are pooled. The evidence overall again suggests that disparities in default rates across lenders are not due to chance alone.

Comparisons across years are presented in Table 17 (in Appendix A), which is the lender analog to Table 14 for tracts. We see that a very small percentage of lenders are identified as "high-

[29] We require that a lender have an expected number of defaults of at least 5 in order to be included individually in the chi-square calculation.

TABLE 16

EXACT PROBABILITY CALCULATIONS ASSUMING THAT LENDER DEFAULT PROBABILITY EQUALS MSA DEFAULT RATE FOR EACH LENDER

PERCENTAGE OF LENDERS BY PROBABILITY OF OUTCOME , BY MSA

PANEL F: 1992 AND 1994 ORIGINATIONS, UNCURED DELINQUENCIES AT TWO YEARS

MSA NAME	TOTAL LENDERS	PROBABILITY OF OUTCOME		
		5% OR LESS	5% TO 10%	>10%
ATLANTA, GA MSA	276	7.25%	4.71%	88.04%
BALTIMORE, MD PMSA	200	5.00	2.50	92.50
CHICAGO, IL PMSA	272	6.25	1.84	91.91
DALLAS, TX PMSA	231	7.36	2.16	90.48
DENVER, CO PMSA	234	5.56	1.71	92.74
DETROIT, MI PMSA	155	6.45	0.00	93.55
FORT LAUDERDALE, FL PMSA	231	3.90	3.46	92.64
FORT WORTH-ARLINGTON, TX PMSA	193	9.33	3.11	87.56
HOUSTON, TX PMSA	156	3.85	1.92	94.23
LOS ANGELES-LONG BEACH, CA PMSA	402	5.47	1.99	92.54
MEMPHIS, TN-AR-MS MSA	99	7.07	2.02	90.91
MIAMI, FL PMSA	242	6.20	2.48	91.32
MINNEAPOLIS-ST PAUL, MN-WI MSA	179	6.15	5.03	88.83
NORFOLK-VIRGINIA BEACH-NEWPORT	121	5.79	2.48	91.74
ORLANDO, FL MSA	162	3.09	4.94	91.98
PHILADELPHIA, PA-NJ PMSA	156	5.13	3.21	91.67
PHOENIX-MESA, AZ MSA	177	4.52	3.39	92.09
RIVERSIDE-SAN BERNARDINO, CA PMSA	451	5.10	3.10	91.80
SACRAMENTO, CA PMSA	195	3.08	4.10	92.82
ST. LOUIS, MO-IL MSA	106	3.77	6.60	89.62
TAMPA-ST PETERSBURG-CLEARWATER	177	3.95	0.56	95.48
WASHINGTON, DC-MD-VA-WV, PMSA	257	7.39	1.95	90.66

default lenders" for both years by our current standard, again suggesting that a portion of what is picked up here is transitory and thus less important.

- **The initial statistical work for lenders suggests that default rates for many lenders are too high to be plausibly explained by chance alone. The identities of these lenders vary with the year of origination, casting doubt on the importance of the differences in default rates.**

4.3. Comparisons with NTIC Designations

The exact probability calculations in the preceding sections identify tracts and lenders for which there is a low probability (5 percent or less) that chance alone could have resulted in as many or more defaults as that observed; we label these as high-default tracts and lenders. It is of interest to compare the tracts and lenders identified in this way with those that would be picked out by the methodology utilized in the NTIC study.[30] Recall that the NTIC methodology identifies tracts as "high-default" tracts if the tract default rate is at least 1.5 times the MSA rate, and it identifies the "10 worst lenders" in each MSA as the 10 lenders with the largest number of defaults.

To preview the findings, when applied to the data on tracts with over 30 loans, the method used by NTIC selects far more tracts as high-default than does the statistical procedure employed here. On balance, the tendency of the NTIC method to attribute too much meaning to default rates in tracts with few loans more than offsets the tendency of the NTIC method to require too high a default rate in tracts with very large loan volume. The NTIC method of identifying poorly performing lenders from among those that originate more than 30 loans identifies about the same number of lenders as does the statistical procedure, but the bias in the NTIC method towards selecting high-volume lenders yields a different assortment of lenders.

To simplify the exposition, we focus the discussion in this section on a single default definition: uncured delinquencies at two years. As usual, Appendix A contains the corresponding analyses for all three default definitions.

To proceed with the evidence on tracts, Table 18 compares the tracts identified as high-default according to the two different standards. Panel B is presented below; the remaining panels of Table 18 are in Appendix A. For purposes of this comparison, we restrict the selection of tracts to those with more than 30 loans in the two origination years combined, and we use exact probability calculations, on the one hand, and the NTIC method on the other, applied to the two years of origination data together. The two columns of each cross tabulation give the classification of a tract according to the statistical methods used in this study, while the two rows give the classification according to the methods utilized in the NTIC study. Each cell of a cross tabulation contains three numbers. The top number is the count for the cell; the second is the row percentage (the cell count divided by the total count for the two cells in the row); and the bottom number is the column percentage (the cell count divided by the total for the two cells in the column). We see in

---

[30] We emphasize that while we use the NTIC methodology, we do not use the data used by NTIC, and thus our results do not necessarily correspond to what would be found in the data actually used in the NTIC study.

TABLE 18

CROSS TABULATION OF HIGH DEFAULT TRACTS AS IDENTIFIED IN THIS STUDY VERSUS HIGH DEFAULT TRACTS AS IDENTIFIED USING NTIC METHODOLOGY*

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

		This Study		
		Non-High Default	High Default	Total
NTIC METHODOLOGY	Non-High Default	4368	6	4374
		99.86	0.14	100
		81.68	1.44	75.87
	High Default	980	411	1391
		70.45	29.55	100
		18.32	98.56	24.13
	Total	5348	417	5765
		92.77	7.23	100
		100	100	100

*Restricted to tracts with more than 30 loans.

the cell in the upper left of Panel B, for example, that when defaults are defined as claims at two years, 4,368 tracts are labeled as non-high-default tracts under both methods, that this number of tracts represents 99.86 percent of the 4,374 tracts labeled as non-high-default tracts using the NTIC methodology, and that the tracts in this cell are 81.68 percent of the 5,348 tracts labeled as non-high-default tracts in this study.

An examination of Table 18 reveals several interesting features. First, almost all tracts classified as high-default in this study are also classified as high-default using the NTIC methodology. Second, the NTIC methodology classifies tracts as high-default overzealously: about 70 percent of the tracts labeled as high-default in the NTIC study are labeled as non-high-default in this study, *i.e.,* do not pass a standard statistical test at conventional levels. These tendencies are reflected in the overall rates; this study identifies 7 percent of the tracts as high-default tracts while the NTIC methodology labels about 24 percent as high-default.

Table 19 (in Appendix A) looks at the same phenomenon in a different metric. Instead of counting tracts according to their classification under alternative labeling schemes, Table 19 counts loans in these tracts. Thus, Table 19 is like Table 18 except that each tract is weighted in accordance with the number of loans in that tract. The message from Table 19, however, is the same as that in Table 18.

Turning to lenders, Table 20 compares lenders identified as high-default or non-high-default in this study to lender representation among the "10 worst lenders," *i.e.,* those with the highest default volume in the MSA, the method used by NTIC. Panel B, presented below, shows that the 10 lenders with the greatest default volume do not appear to match at all well with those lenders with default rates that are significantly different from the MSA rate. (Panels A and C of Table 20 are in Appendix A.) The list of the 10 worst lenders misses 60 percent of the lenders identified as high-default in this study. Moreover, about 63 percent of the list of 10 high-default-volume lenders are labeled as non-high-default using the methods employed in this study. Both methods single out about 10 percent of the lenders, but the methods disagree strongly over which lenders should be identified as problem lenders.

Table 21 (in Appendix A) repeats the analysis of Table 20, except that now each lender is weighted according to its loan volume. Although results here are qualitatively similar to those in Table 20, they show, not surprisingly, that the "10 worst lender" method picks out large lenders, so that while only about 10 percent of lenders are selected by this method (Table 20, Panel B, above), these lenders account for about 39 percent of loans (Table 21, Panel B, in Appendix A). This size bias also explains why only 21 percent of the loans made by lenders classified as high-default in this study are made by lenders that are not in the "10 worst lenders" list, yet 60 percent of the lenders identified as high-default in this study are not in the "10 worst" list (Table 20, Panel B, above).

Tables 22 and 23 (in Appendix A) repeat the analysis in Tables 20 and 21 using an alternative way of picking out problem lenders. Here we use lender identification numbers, provided by HUD, to identify the 10 worst lenders in each city that are singled out by the NTIC study. For the ten MSAs analyzed in this study that also appear in the NTIC study,[31] we produce

---

[31] These MSAs are Atlanta, Baltimore, Chicago, Denver, Detroit, Los Angeles-Long Beach, Minneapolis-St.Paul, Philadelphia, St.Louis, and Tampa-St. Petersburg.

TABLE 20

CROSS TABULATION OF HIGH DEFAULT LENDERS AS IDENTIFIED IN THIS STUDY VERSUS TEN LENDERS WITH HIGHEST DEFAULT VOLUME*

1992 AND 1994 ORIGINATIONS

PANEL B: UNCURED DELINQUENCIES AT TWO YEARS

		This Study		
		Non-High Default	High Default	Total
Default Volume	Non-High Default Volume	1851	123	1974
		93.77	6.23	100
		93.02	60.29	89.97
	High Default Volume	139	81	220
		63.18	36.82	100
		6.98	39.71	10.03
	Total	1990	204	2194
		90.7	9.3	100
		100	100	100

*Restricted to lenders with more than 30 loans.

a cross-tabulation of whether the lender is a high-default lender as identified in this study, by whether the lender is on the list of the "10 worst lenders" in the NTIC study. As seen in Table 22 (in Appendix A), the findings are as might have been anticipated from Table 20: there is disagreement in both directions, but the NTIC method is generally too quick to label a lender as high-default. Table 23 (in Appendix A), which provides a parallel analysis in which each lender is weighted by loan volume, tells the same story as Table 22.

- **Tables 18 through 23 show that the methods employed by NTIC can be either too stringent or too lenient in identifying problem tracts and problem lenders. Whether leniency or stringency dominates overall depends on the distribution of loan volumes and on MSA default rates. In the data used in this study, the NTIC methods generally identify tracts as high-default tracts overzealously and label the wrong set of lenders as high-default.**

It is particularly easy to use data presented in the NTIC report to illustrate the possibility of overzealousness in identifying high-default tracts in their own data, *i.e.,* labeling individual tracts as high-default that may easily owe their high default rates to chance. The NTIC study reports the MSA default rates and the average loan totals in the tracts they single out as high-default in each city. Utilizing their criterion of high-default (having a default rate that is at least 1.5 times the MSA rate), we can compute probability of obtaining the minimum number of defaults, or more, that would result in a label of high-default status in a tract of average size, under the assumption that the tract-level default probability is actually the same as that at the MSA level.[32] These probabilities are reported in the last column of Table 24 (in Appendix A). Hence, this column shows the probability that a tract with average loan volume would be labeled as a high-default tract by the NTIC standard even though its defaults occurred randomly at the MSA rate. These probabilities are generally in the 20 percent to 30 percent range, far higher than the probability conventionally applied in statistical work. That is, the NTIC method will in this case result in overzealous labeling of tracts as high-default.

The opposite problem is also a real possibility with the NTIC methodology, especially if applied to data with many loans per tract. That is, with large sample sizes, the requirement that tract default rates be at least 1.5 times the MSA rate can be far too demanding in the sense that tracts in which defaults occur far too frequently to be plausibly generated by chance may escape detection. More generally, a simple decision rule like that employed by NTIC is an inappropriate detection tool, and better methods are readily available.

[32] The minimum number of defaults was rounded up to the next higher integer.

SECTION 5

STATISTICAL ANALYSIS OF DATA ON INDIVIDUAL LOANS

5.1. Analysis of Defaults

The statistical evidence in Section 4 suggests that there are geographic and interlender disparities in default rates that are not due to chance alone. Those analyses, however, make no allowance for other factors that may account for differences in default rates. In this section we continue the investigation by taking a more detailed look at what underlies the default behavior of individual loans. Such an analysis will enable us to see whether the individuals who are located in what have tentatively been identified as high-default tracts, or served by what have tentatively been called high-default lenders, are more likely to default even after controlling for characteristics of the loan and borrower, and we will be able to measure directly any effects on default stemming from presence in a high-default tract or service by a high-default lender.

The main tool for this investigation is a logit analysis[33] of defaults of purchase money loans[34] for each of the 22 MSAs. We perform a separate analysis for each of the three measures of default, and analyses using all three measures are presented in Appendix A, but again the discussion is largely confined to defaults defined as uncured delinquencies at two years. To measure the impacts of interest, we include indicators for high-default tracts and high-default lenders specific to each origination year. That is, to determine the impact of residence in a high-default tract, we include an indicator for whether a 1992 loan is in a high-default tract for 1992 originations (ctin92, ntin92, and n95tin92, for claims at two years, uncured delinquencies at two years, and uncured delinquencies at 12/95, respectively). A separate indicator (ctin94, ntin94, and n95tin94) shows whether a 1994 loan is in a high-default tract for 1994 originations. Similarly, we include indicators for whether a 1992 loan has been originated by a high-default lender (cin92, nin92, and n95in92) and for whether a 1994 loan was made by a high-default lender (cin94, nin94, and n95in94).[35] The determination of "high-default" is made from the analyses in Tables 13 and 16; a tract (lender) is treated as high-default for a particular origination year if the probability of obtaining as many or more defaults is 5 percent or less using data from that origination year alone.[36] We further restrict

---

[33] We use logit, as opposed to linear regression, because of the qualitative, dichotomous nature of the dependent variable. A hazard model would probably be a superior alternative, especially for modeling default behavior over longer intervals of loan duration, though logit is adequate for present purposes.

[34] Some of the variables included in the logit analysis are available only for purchase money loans, which also constitute the vast majority of loans at issue.

[35] The decision to include separate year-specific indicators is an attempt to reduce the importance of changes in tract definitions between the two origination years.

[36] By this criterion, high-default tracts contain 23 percent of defaults under the claims-at-two-years definition, 22 percent of uncured delinquencies at two years, and 25 percent of uncured delinquencies at 12/95. High-default lenders contain about 21 percent of claims at two years, 27 percent of uncured delinquencies at two

this label to tracts and lenders for which the underlying volume of loans in the tract (lender) exceeds 30 in the origination year at issue.[37]

5.1.1. Explanatory Variables

Among the controls is a set of variables that are intended to capture the influence on default of several standard underwriting criteria. These variables include the loan-to-value ratio, the back end ratio (the ratio of housing expenses and other debt payments to income), the front end ratio (the ratio of housing expenses to income), assets after closing, and monthly income. Other controls, such as age, race, number of dependents, etc., which are not recognized in underwriting, but which seem empirically to affect defaults, are included as well.[38] The full list of explanatory factors, other than the high-default tract and high-default lender indicators, is as follows.

ltv:	Loan-to-value ratio (expressed as a spline[39] with a breakpoint at 0.95)
back:	Back end ratio (expressed as a spline with a breakpoint at 0.36)
front:	Front end ratio (a spline with breakpoint at 0.27)
asst:	Assets after closing (entered as a spline with breakpoints at $6,000 and $10,000)
incdiff:	Monthly income (expressed as monthly income minus the MSA average of monthly income)
_94:	Indicator for the 1994 origination year
age:	Age of borrower (a spline with breakpoints at 30 and 40)
less15:	Indicator for loan term of 15 years or less
mtgdiff:	Loan amount (with MIP) expressed as a deviation from the MSA average
intdiff:	Note rate (expressed as deviation from the MSA average, and splined with a breakpoint at 0.7)
sepmale, sglmale:	Indicators for separated borrowers and for single male borrowers
sepfmle, sglfmle:	Indicators for separated borrowers and for single female borrowers
armflag:	Indicator for ARMs
condo:	Indicator for condominiums
firstime:	Indicator for first-time buyer
black:	Indicator for black borrower
hispan:	Indicator for Hispanic borrower

years, and 26 percent of uncured delinquencies at 12/95.

37 Rerunning the analysis with a cutoff of 10 loans, rather than 30, appeared to give qualitatively similar results.

38 See Neal (1989) and Quercia and Stegman (1992) for useful summaries of the default literature.

39 Breakpoints in splines were determined by casual observation of bivariate plots of means of explanatory variables against default rates.

cnincdif:	Tract income divided by MSA median (from Census files[40])
blkcen:	Fraction of tract population that is black (from Census files)
hspcen:	Fraction of population that is Hispanic (from Census files)
unempcen:	Tract unemployment rate in 1990 (from Census files)
fhaorig:	FHA originations divided by total originations (from HMDA files[41])
cnvadeny:	Conventional denials divided by conventional applications (from HMDA files)
hasasset:	Indicator variable for the presence of positive assets
hascen:	Indicator variable for the ability to match the tract to Census data
hashum:	Indicator variable for the ability to match the tract to HMDA data

In addition to the MSA-specific analyses, for expository purposes we run pooled versions of essentially identical logits in which data from all 22 MSAs are included in a single estimation procedure. In contrast to the MSA-specific logits, the pooled logits include indicators for the particular MSA, as well as two variables that are measured at the MSA level --- "avgrate," the average MSA unemployment rates from origination through mid-1997 (from the BLS), and "house," percentage MSA house price growth from origination through the first quarter of 1997 (from the Freddie Mac Repeat Sales Index).[42] Because the latter two variables are obtained at the MSA level, they exhibit variation within an MSA only because of differences in dates of loan origination; these variables have been excluded from the MSA-specific analyses. They have, however, been included to help explain differences across MSAs in the pooled analyses.

5.1.2. Logit Estimates of Pooled Data

Because the sheer volume of output from the logits on the 22 individual MSAs is so large, we present only the full estimates from the pooled model;[43] the MSA-specific analyses are presented in Appendix B (available from HUD). These pooled logit estimates are presented in Table 25 (in Appendix A). The first column of numbers in each of the three panels presents the coefficient estimate,[44] the second is the standard error, the third is the asymptotic normal statistic (z), the fourth gives the significance level (probability of obtaining results this extreme if the true effect is zero,

[40] Those loans in tracts that cannot be matched to Census data are assigned values of zero for all Census-derived variables, and an indicator variable, "hascen," is assigned a value of one.

[41] Those loans in tracts that cannot be matched to HMDA data are assigned values of zero for all HMDA-derived variables, and an indicator variable, "hashum," is assigned a value of one.

[42] For the model using defaults defined as uncured delinquencies at 12/95, we also include a variable "month" that measures potential loan duration, the number of months from origination to December 1995.

[43] Coefficient estimates for the MSA indicators are not of interest for current purposes and are suppressed.

[44] The coefficient estimates in Table 25 measure estimated impacts on the logit index function and are therefore difficult to interpret directly.

based on a two-tailed test), and the final column is a pair of numbers that constitute a 95-percent confidence interval. As can be seen in the three panels of Table 25, the effects of high-default tracts and high-default lenders appear to be significantly different from zero, even after controlling for a variety of default-related factors. The findings also show that the control factors generally do matter, typically in the anticipated direction.[45] Notice also that the qualitative effects of these default-related factors are in line with what we have seen in Tables 6 and 11. For example, according to Table 25, higher LTVs are associated with higher default probabilities, while according to Tables 6 and 11, tracts and lenders with higher default rates have more substantial fractions of borrowers with high LTVs. Similarly, according to the findings in Table 25, higher default probabilities among individual loans are associated with lower asset levels; with first time buyers, black borrowers, and non-Hispanic borrowers; with lower tract income relative to the MSA; and with higher conventional denial rates in the tract. All of these factors, when measured in the various tract or lender default-rate categories, are also associated with higher aggregate default rates, according to the data presented above in Tables 6 and 11.

- **These findings reinforce the notion that lenders and tracts have high default rates partly because their loans tend to be riskier. Clearly, numerous factors, only some of which can or should be considered in underwriting, have effects on default.**

5.1.3. Logit Results for Individual MSAs

Despite the fact that controlling for a variety of default-related factors still leaves area and lender impacts, their magnitudes are considerably reduced. To see this, we turn to a summary of the results from logits estimated over each of the individual MSAs; all of the individual MSA-level logits are presented in Appendix B. This summary, presented in the three panels of Table 26, shows the estimated impact of high-default tracts and lenders in "raw" form, *i.e.*, before we control for other factors,[46] and in "adjusted" form, *i.e.*, after we control for other factors. Panel B is given below and is the main focus of the discussion. (Panels A and C are in Appendix A.) The raw and adjusted numbers in this table are expressed as estimated effects on the odds ratio. That is, each number in the table is the estimated multiplicative effect (of a high-default tract or a high-default lender) on the odds of default, where the odds of default are the probability of default divided by the probability of nondefault. Thus, for example, Panel B of Table 26 indicates that for the Washington, DC, PMSA, the raw effect of being in a high-default tract in 1992 was to increase the odds of default by a factor of 5.69. After adjusting for the factors included in the logit, the estimated effect falls to an adjusted impact of 2.72; that is, after adjustment, being in a high-default

[45] Note that variables that are splined are represented by as many coefficients in the logit as there are segments in the spline. Each such coefficient is identified with the basic variable name (*e.g.*, ltv) as the first part of the coefficient name. The first such coefficient name gives the effect in the first segment of the spline, and each succeeding coefficient measures the marginal effect (and its significance level). Thus, the total effect of the variable in any segment of the spline is the sum of the coefficients pertaining to that segment and all previous segments.

[46] Estimates of raw effects were obtained from MSA-specific logit analyses that included only the year-specific high-default indicators and an indicator for 1994 originations.

TABLE 26

Raw and Adjusted Odds Ratios for High-Default Tracts and High-Default Lenders from MSA-Specific Logits

Panel B: Uncured Delinquencies at Two Years

MSA Name	Tracts								Lenders							
	1992				1994				1992				1994			
	raw	Z	adjusted	Z	raw	Z	adjusted	Z	raw	Z	adjusted	Z	raw	Z	adjusted	Z
ATLANTA, GA MSA	5.35	6.66	3.12	4.08	2.92	5.42	1.97	3.11	2.54	5.38	1.63	2.41	2.26	6.85	1.79	4.50
BALTIMORE, MD PMSA	7.70	7.19	4.18	4.79	3.60	6.86	2.09	3.63	2.95	3.66	1.55	1.29	2.08	6.12	1.57	3.42
CHICAGO, IL PMSA	4.91	11.05	2.56	5.96	3.78	13.90	2.06	6.70	1.95	5.54	1.41	2.60	2.06	9.96	1.59	5.78
DALLAS, TX PMSA	4.15	6.53	2.70	3.77	5.48	8.64	3.08	4.95	3.62	7.52	2.79	4.82	2.52	5.28	1.73	2.87
DENVER, CO PMSA	14.89	8.90	8.41	6.27	4.95	6.89	3.95	5.44	5.85	2.86	6.46	2.90	2.18	4.28	2.03	3.64
DETROIT, MI PMSA	4.67	6.10	2.74	3.62	3.44	8.29	1.71	3.14	3.88	6.16	1.30	1.06	5.02	12.05	2.15	4.93
FORT LAUDERDALE, FL PMSA	5.08	2.58	2.41	1.13	1.91	2.17	1.89	2.02	2.15	1.82	1.95	1.52	2.31	3.80	2.06	2.97
FORT WORTH-ARLINGTON, TX PMSA	4.09	2.80	2.52	1.45	2.19	2.84	0.72	-0.44	4.16	4.63	2.50	2.37	5.84	7.27	3.22	4.04
HOUSTON, TX PMSA	0.90	-0.14			2.26	3.45	4.40	3.87	2.38	3.31	1.94	2.04	1.58	1.96	1.02	0.07
LOS ANGELES-LONG BEACH, CA PMSA	1.63	2.11	1.78	2.35	2.35	8.56	1.99	5.77	2.98	6.60	2.78	5.53	2.17	6.29	2.10	5.57
MEMPHIS, TN-AR-MS MSA	3.58	4.93	1.91	2.18	3.84	8.54	2.17	4.34	2.47	3.78	1.82	2.26	1.45	1.67	1.42	1.38
MIAMI, FL PMSA	3.45	3.18	2.72	2.44	2.90	7.34	2.91	5.38	5.10	6.49	3.75	4.89	2.81	8.07	2.10	5.05
MINNEAPOLIS-ST PAUL, MN-WI MSA	11.39	9.34	7.32	7.04	4.40	6.33	3.77	5.35	3.58	4.94	3.26	4.17	1.87	3.43	1.82	3.02
ORLANDO, FL MSA	1.38	0.32	1.67	0.49	4.28	6.89	2.45	3.55	2.79	2.50	2.70	2.32	1.84	3.80	1.29	1.36
PHILADELPHIA, PA-NJ PMSA					2.05	5.63	1.46	1.63	5.06	4.02	3.55	2.98	1.68	4.43	1.41	2.59
PHOENIX-MESA, AZ MSA	4.69	6.39	3.69	4.95	3.83	7.42	3.35	6.26	1.72	2.87	1.58	2.20	1.69	3.33	1.39	1.92
RIVERSIDE-SAN BERNARDINO, CA PMSA	2.64	7.95	2.23	5.43	2.19	7.06	1.99	5.76	1.96	5.12	1.88	4.23	1.52	4.83	1.44	3.80
SACRAMENTO, CA PMSA					4.03	4.84	2.68	2.90	3.69	1.75	2.39	0.81	3.89	3.54	4.23	3.36
ST. LOUIS, MO-IL MSA	6.65	4.30	3.19	2.45	2.41	4.85	2.04	2.10	2.90	2.03	2.17	1.42	2.66	4.14	1.99	2.66
TAMPA-ST PETERSBURG-CLEARWATER					1.79	2.65	2.74	3.16	3.38	4.71	2.82	3.62	2.34	5.17	1.75	3.05
WASHINGTON, DC-MD-VA-WV, PMSA	5.69	6.46	2.72	3.00	2.48	9.07	2.63	6.63	2.23	4.70	1.86	3.27	2.25	9.03	1.73	5.51

tract is estimated to increase the odds of default by a factor of 2.72, rather than 5.69. The reduction in the estimated odds ratio is impressive, but the estimated impact after adjustment remains fairly high. In addition, we present the asymptotic normal statistic (z) for each coefficient estimate.[47]

Comparing the raw with the adjusted effects in Table 26, we see that for the vast majority of MSAs, there is a substantial decline in the estimated impact of high-default tracts and lenders. The aforementioned change in the estimated impact for high-default tracts in 1992 in the Washington, DC, PMSA is but one example. There are some cases in which the estimated impacts rise after introducing controls, particularly in Panel A, but these cases are relatively rare.[48] In addition, there are a number of cases in which the adjusted effects are no longer estimated to be significantly different from zero.[49] Thus, the fractions of tracts and lenders that are labeled as high-default --- fractions that were already substantially below the fractions yielded by the NTIC methodology (in the case of tracts) --- are thus further reduced, and the estimated impacts of high-default tracts and lenders are generally reduced as well.

- **To conclude, we find that controlling for a variety of default-related factors usually reduces the estimated impact on default of residence in a high-default tract or origination by a high-default lender. In most MSAs, however, estimated impacts are still significantly different from zero.**

5.1.4. Possible Reasons for Area and Lender Effects

Even after allowing for the influence of a wide variety of factors, there are typically still significant effects of high-default tracts and lenders. Although underwriting practice and lender servicing could be the problem, there are numerous other possibilities that are also plausible and which deserve mention. First, important underwriting factors, particularly all aspects of the borrower's credit history have been omitted from this analysis simply because we lack such data. Although we cannot tell whether this omission is the only factor leading to the appearance of tract and lender effects, it likely is a very important contributor.

Second, even if all underwriting guidelines are followed perfectly, the uneven distribution of house price growth, unemployment, etc., will likely lead to pockets of defaults. We lack detailed local information on such factors at the tract level; the information we do have is measured at the MSA level. Tract-to-tract differences are likely and might help explain intertract and interlender differences in default rates. Even such factors as illness, death of the borrower, and divorce, which

---

[47] The standard we use most other places in this paper is a significance level of 0.05 in a one-tailed test, for which the corresponding value of z is 1.645. For a two-tailed test at a significance level of 0.05, the value of z is 1.96.

[48] We may in some instances have poor estimates of the effect of other factors, leading perhaps to greater impacts being attributed to lender and tract.

[49] There are also some cases in which the raw effect is estimated to be insignificantly different from zero. This anomalous finding may reflect (a) our use of "good" data on purchase money loans for the logit, in contrast to the use of many more loans for the exact probability calculations, and (b) the use of a different probability model for the logit than for the exact probability calculations.

may help precipitate default, can display geographic differences that, while temporary, could lead to corresponding temporary differences in default rates.

Third, the data are surely not error free. Virtually all of these series are error-ridden measures of what they purport to measure. In many cases, even a perfectly measured variable would be only a readily available proxy for the unavailable variable that we would prefer to obtain. Measurement errors in explanatory variables generally reduce the ability of these factors to explain defaults, often causing other related variables to appear significant. Suppose, for example, that perfectly measured LTV is positively related to the incidence of default and also varies across tracts and lenders. We may find that an imperfectly measured LTV variable will fail to pick up all intertract and interlender variation in defaults arising from variation in true LTV, with the result that tract and lender effects fail to vanish even when controlling for measured LTV.

Fourth, the ability of the statistical analysis to explain defaults rests on assumptions about the way in which each of the included factors affects defaults. Failure to represent properly the way in which these factors operate mathematically could also lead to incomplete adjustment for these controls and an appearance that particular tracts or lenders have inexplicably high default rates.

Given the substantial declines in the estimated impacts of high-default tracts and lenders when we introduce other controls --- many of which may be more closely related to longer term defaults than to the early defaults examined here --- it is not inconceivable that a full set of controls would reduce estimated impacts to essentially zero. There is, however, no way of telling in advance of actually performing the operation. Thus, while the evidence contained herein indicates the existence of high-default tracts and lenders in at least some MSAs, this conclusion should be strongly tempered by the realization that these findings might well vanish if we could better measure all appropriate determinants of default.

It is possible, of course, that the appearance of high-default tracts or lenders arises because of poor underwriting and overly eager foreclosure policies by lending institutions. The findings above do not imply that underwriting guidelines are being followed by individual lenders or that guidelines are being followed in all areas, nor do they imply the contrary. That is, default related factors will have effects whether or not underwriting excludes those deemed to be "poor risks." Indeed, even if a set of default-related factors were to reduce all estimated tract and lender effects to zero, it would not follow that underwriting guidelines were or were not being followed. Unfortunately, direct testing of whether guidelines are being followed would require relatively error-free data on all underwriting factors, as well as a way to quantify correctly all underwriter judgement. These tasks are beyond the ability of FHA data to deliver.

- **Lack of data on credit history of borrowers, together with a variety of statistical difficulties, may in part explain why some tracts and lenders appear to affect default probabilities adversely. The FHA data do not permit us to tell whether underwriting criteria are or are not followed, and whether differences in default rates are in any way traceable to lax underwriting.**

5.1.5. Additional Comparisons with the NTIC Findings

In Section 4.3 we compared the numbers of high volume tracts and lenders (those having greater than 30 loans) that this study identifies as high-default entities with high volume tracts or lenders identified as high-default using the NTIC methodology. Both the statistical methodology and the NTIC methodology employed to identify high-default entities are based on comparisons of each tract or lender individually with the MSA as a whole. There are no corrections for differences among the loans or borrowers serviced in the tracts. In addition, there is no allowance for the fact that the presence of high-default tracts or lenders (or low default tracts or lenders, for that matter) in the MSA affects the default rate in the pool of all MSA loans against which comparisons are made. The statistical work in Section 5 adjusts for a host of default-related factors in evaluating whether the tracts and lenders labeled as high-default by the simple statistical methodology have, as groups, significant effects on default probabilities of individual loans.

In this section we ask whether any of the high volume tracts or lenders that the NTIC methodology labels as high-default, but which are not labeled as high-default by the simple statistical methodology utilized in Section 4, have significant effects on default probabilities. Because of the burden of carrying out this exercise for each MSA, we use only the Chicago MSA as an example, though we cannot, of course, be certain whether this example is representative. In addition, we perform the exercise only for the default definition that uses uncured delinquencies at two years. To perform this exercise, we again use purchase money loans in the Chicago MSA to estimate a logit model like that described above; in this case, however, we replace the indicators for the high-default tracts and lenders as identified with the statistical methodology with indicators for each of the high-default tracts and lenders identified using the NTIC methodology that were not also identified as high-default by the statistical methodology.[50] The result is the inclusion of 12 origination-year-by-lender indicator variables and 63 origination-year-by-tract indicator variables to pick up the combinations of lenders and years, and the combinations of tracts and years, labeled as high-default by the NTIC methodology only. Upon estimating this logit, we find that for 9 of the 12 lender indicators, the estimated effects are of the "wrong" sign, *i.e.,* the estimates indicate that loans from these lenders have *lower* default probabilities for that origination year, other things the same, and the other 3 indicators are of the "right" sign but are not significantly different from zero (at even a 10-percent level in a two-tailed test). For tracts, we find that 24 indicators are of the "wrong" sign, another 21 are of the "right" sign but statistically insignificant, and 18 are of the "right" sign and significantly different from zero at the 10-percent level or better (in a two-tailed test).[51]

---

[50] More precisely, to ensure that the results are not biased by using one sample for the logit and a slightly different sample to identify the high-default tracts and lenders, we rerun the analysis to identify high-default tracts and lenders on the sample of Chicago purchase money loans for the two years separately (for identifying high-default entities under the statistical method) or for the two years together (using the NTIC method). We identify tracts and lenders as high-default for the logit only if the tract or lender has more than 30 loans in the origination year of the individual loan observation and is singled out by the NTIC method and not the statistical method for that particular origination year.

[51] Introducing four additional indicators for the year-specific groups of high-default tracts and high-default lenders identified in this study, we get qualitatively similar results for the coefficients on tracts and lenders that are held to be high-default in the NTIC study only. Among the individual tracts identified as high-default under the NTIC method alone, we find 18 have the "wrong" sign, 25 have the "right" sign but are statistically insignificant, and 20 are statistically significant with the "right" sign. Among lenders identified as high-default under the NTIC method only, we find that 4 have the "wrong" sign, 5 have the "right" sign but are statistically

The fact that coefficients for 18 of these tracts show up with significant effects, whereas none showed a statistically significant difference in the simpler statistical tests based on the binomial, deserves explanation. There are at least two reasons for these seemingly inconsistent findings. First, in the logit we control for the influence of a variety of default related characteristics of the loan and borrower when asking whether the tract has an effect; in the simple statistical tests there are no such controls. Second, in estimating the effect of each tract and lender for which an indicator is entered, the logit removes the effect of all other tracts and lenders for which separate indicators are entered; in contrast, the simpler statistical procedures compare each tract or lender against the MSA as a whole. To illustrate the importance of latter point, suppose there are four tracts, A, B, C, and D. The simple statistical methods used in Section 4 decide whether to label tract A as high-default by comparing the default rate in A to that in the pool composed of A, B, C, and D. In contrast, abstracting from other default related factors, a logit analysis that introduces indicators for, say, tracts A and B, implicitly compares the default rate of loans in tract A to loans in the pool composed of loans from C and D only. Removing tracts B and A from the comparison group can thus give different results than if these loans were included.

One should not conclude that this finding justifies the nonstatistical methods employed by NTIC; as noted the selection of lenders and tracts singled out by NTIC's methods turned up many with estimated effects in the "wrong" direction. The lesson here is that there are interdependencies in estimating logits, and estimated effects for each tract and lender may be sensitive to the inclusion of indicators for other tracts and lenders. Indeed, by searching for, and including indicators for, "low default tracts" and "low default lenders" one may find that some of the tracts previously estimated to be high-default are no longer classified as high-default. Labeling a tract or lender as high-default inevitably depends on the composition of the comparison group. To help prevent overlooking possible problem areas, it may be somewhat safer to start by statistically identifying overly inclusive groups of high-default tracts and lenders, letting more detailed statistical analysis determine which ones survive after correction for observable differences in borrowers and loans.[52]

5.2. A Tentative Analysis of Lender Responses to Delinquencies

As noted, the FHA data accessed here exclude potentially crucial ingredients in the underwriting decision, thus making it impossible to account for a full complement of default-related factors and similarly making it impossible to ascertain directly whether underwriting guidelines are being followed. The existing FHA delinquency data do, however, permit at least a crude analysis of two aspects of lender servicing behavior: the frequency with which alternatives to foreclosure are offered and the speed with which lenders intervene in a delinquency. Both of these seem to be of concern in the NTIC study, though no real evidence is offered there on either dimension of lender behavior. We pursue this brief empirical study with some misgivings; this analysis relies heavily

insignificant, and 3 have the "right" sign and are statistically significant.

[52] Alternatively, one could select problem areas by comparing actual default rates to rates predicted on the basis of a model, estimated nationally or locally, that takes account of characteristics of loans and borrowers. The difficulty with the latter method is in appropriately allowing for current conditions in local housing markets that may make them behave unusually well or unusually poorly.

on the completeness and accuracy of the delinquency data that have already been called into question. We thus proceed provisionally and with even more than the usual dose of *caveat emptor*.

The approach is as follows. We extract each occurrence of a 90-day delinquency in all of the data on the 1992 and 1994 originations that underlie this study, and we then classify the very next event into one of three categories: (1) censored by the closure of the data series (*i.e.*, no additional events are recorded), (2) cure instituted by the borrower, or (3) action taken by the lender, which could include a move to foreclose or an offer of an alternative to foreclosure (defined to include an offer of forbearance, repayment plan, modification, or assignment[53]). We then ask whether actions taken by the lender occur more quickly for lenders that we have identified as high-default than for non-high-default lenders.[54] For this calculation we treat both censored-by-close-of-data and cures as censored events, for both of these actions preclude our observing what the lender would have done in the absence of these intervening events. Using Cox regression to estimate a simple hazard model of the instantaneous probability of lender intervention, we find that high-default lenders intervene more quickly ($z=2.685$) when we use the claims-at-two-years definition of default, less quickly when we use the delinquencies-at-two-years definition of default ($z=-1.86$), and less quickly (though not significantly so) when we use the uncured-delinquencies-at-12/95 definition of default. The results regarding the speed of lender intervention are thus ambiguous.[55]

Because the lender interventions analyzed above include both movements to foreclose (which some would consider "bad") as well as offers of alternatives to foreclosures (which might be considered "good"), we further ask whether high-default and non-high-default lenders differ in the frequency with which one or the other is offered, given that some lender action is taken. Table 27 (in Appendix A) presents the appropriate cross tabulations. (Contents of cells follow the same format as the cross tabulations presented in Tables 18 through 23.) Notice that only about 8 to 10 percent of the lender actions fall into the "good" category, and that we again have mixed results. When the first definition of default is used (Panel A), high-default lenders offer alternatives to foreclosure more frequently than do non-high-default lenders, and the differences are significant at conventional levels. In Panel B high-default lenders are observed to move to foreclosure more frequently, while in Panel C, the opposite is true. In neither of the latter two panels do we see differences that are statistically significant at conventional levels, however.

- **Putting aside concerns about the data used in this exercise, the evidence in this section points to no consistent and important difference in the speed with which high-default lenders act on a delinquency, nor do we see any substantial evidence of a bias towards foreclosure among high-default lenders responding to delinquencies.**

---

[53] Inclusion of assignment could be problematic since borrowers could gain entry into the program even without lender support.

[54] For all of this section, we identify high-default lenders using the exact probability calculations in Table 16 for the two origination years together, and we restrict attention to lenders with at least 30 loans in the two years together.

[55] Reformulating the hazard model by reclassifying as censored all offers of alternatives to foreclosure (thus treating only movements to foreclose as uncensored) also yields ambiguous findings. None of the high-default lender effects were significantly different from zero at conventional levels.

SECTION 6

A SUMMARY OF THE METHODOLOGY AND GENERAL FINDINGS

6.1. Choosing Time Periods, Default Definitions, and Areas

Here we review the methods employed in this study, as well as a few of the general findings. The reader is referred to the summary at the beginning of the paper for a statement of more specific and detailed findings and conclusions.

The presumed main purpose of an analysis like that reported here is to identify neighborhoods and lenders for which high default rates are indicative of some failure. Although such failure could include poor servicing performance by lenders, the possibility of poor underwriting practice is a major focus. Poor underwriting practice could take two forms: either guidelines promulgated by FHA are not being followed, or the FHA underwriting criteria are themselves defective. Because the major focus is on underwriting, it is sensible to restrict the investigation to evidence that will most clearly indicate underwriting deficiencies. In our view, a focus on default activity in the early years reflects most clearly on underwriting, in part because many of the factors that can be measured at loan qualification can and do change over time. Even though autocorrelation tends to make factors (*e.g.*, income) measured at loan qualification fairly predictive for some time afterward, even these autocorrelations are not strong enough to make for excellent predictive power many years after underwriting takes place. In addition, a focus on relatively short loan durations will still permit identification of other difficulties; for example, servicing problems seem likely to show up in default experience at all loan durations.

As noted earlier, we do not deny that default activity far later in the life of the loan could still be influenced by underwriting policy. A dramatic decline in the acceptable level of the loan-to-value ratio, for example, could reduce default activity many years after loan origination. It is simply that many of the effects of underwriting factors become more tenuous as loan duration rises. The current study operationalizes this idea by focusing on two measures of default that measure loan status two years following loan origination, and a third that measures loan status at up to four years.

There is a related question of how to define a default. The operating assumption in this paper is that while there are some costs to delinquencies, foreclosures are the major events to be avoided, and the purpose of the analysis is ultimately to help reduce foreclosure probabilities. A broad definition that includes as defaults all loans in delinquency status will pick up many loans that will never enter foreclosure, and this definition is thus too broad. At the other extreme, restricting the definition of default to loans that have gone to claim will miss many loans that are in the earlier stages of a foreclosure process that is certain to be completed. Moreover, the speed with which delinquencies make their way to claim depends in part on state foreclosure practices and laws. Including as defaults those delinquencies that have not been observed to cure for at least several months after becoming 90-days delinquent seems to be a reasonable middle ground. Including these delinquencies-in-progress is likely to pick up most delinquencies that will terminate in foreclosure at the conclusion of the existing spell, while excluding many loans with delinquency spells that are likely to end in cure. Two of the measures examined in this paper follow the latter strategy, while the other uses only claims to measure default activity.

Having decided on a definition of default and a time at which to measure delinquency, the next task is to look at default activity so defined within an appropriate geographic area and within an appropriate set of origination years. Time of origination is relevant in part because the underwriting criteria or their enforcement --- *i.e.,* the objects at issue --- may themselves change, but also because changes in the economic environment may alter the default rates against which performance is to be measured. For similar reasons, it is desirable to focus an investigation of geographic dispersion in default rates on relatively homogeneous areas so that problem areas are not hidden by aggregation with areas that are not suffering from high default rates. The current study assumes that census tracts are homogeneous neighborhoods and that two origination years (1992 and 1994) which are separated by one year are sufficiently homogenous. We find, however, that there are substantial differences in the tracts and lenders that are identified as high-default in the two origination years used in this study.

In building a sample by using small neighborhoods or narrow intervals of loan origination, one trades off sample size for homogeneity, and there are several costs in reducing sample sizes that should be considered. First, small samples make it difficult to detect default rates that are substantially higher than the standard against which one is measuring. Suppose, for example, the MSA default rate is taken as the standard for comparison, and the MSA default rate over the relevant time interval is, say, 5 percent. Suppose further that we agree to use a significance level of 5 percent in our tests. That is, we will reject the notion that the true, underlying default probability in an area is the same as the MSA rate (and thus we will label the area as a high-default area) only if the probability of finding at least as many defaults as what we actually observe is calculated to be 5 percent or less when the area default probability is the same as the MSA default rate. If a particular area has only five loans over this same period, our decision rule would lead us to assign high-default status to the area if we find two or more defaults among the five loans. Suppose, however, that the true, underlying probability of default in the area were 20 percent. The probability of finding one or fewer defaults among the five loans, and thus incorrectly failing to label the area as a high-default area, is about 73 percent. There is thus an excellent chance that a small area where the true underlying probability of default greatly exceeds the underlying MSA rate will escape detection.

Another advantage in large sample sizes is that they increase the accuracy of tests based on large sample approximations. As seen in our discussion of the chi-square tests, the combination of small samples and rare events can make some standard tests virtually unusable.

An unrelated difficulty with utilizing areas or lenders with small loan volume is that the potential gains from reductions in the default rate will tend to be small as well. There are likely to be costs in conducting a statistical investigation and in instituting and monitoring corrective action when called for, and these resources are better spent where the prospective gains are largest, which will generally mean dealing with larger areas or lenders, or only those small-volume areas or lenders with especially high default rates.

6.2. Statistical Analysis of Default Probabilities

Having identified relatively homogeneous observational units --- areas or lenders --- the next simple step is to make an initial statistical exploration to compare performance to the benchmark. A useful benchmark is a rate calculated over a larger area (such as the MSA) in the same time period, under the assumption that an MSA-wide average recognizes that idiosyncratic events may affect the area as a whole and cause it to have a different experience than other MSAs. One does not want to mix in these effects with any due to policies of individual lenders. At the same time, the use of MSA rates as a benchmark in low-default MSAs means that some tracts and lenders labeled as high-default may have default rates that are low enough to be considered unimportant from a practical standpoint. A related implicit assumption is that within-MSA comparisons are not distorted by differences in, say, the average quality of lenders across MSAs. For example, a comparison of the default rate of each lender to that in the MSA as a whole assumes that the default rate in the MSA as a whole is not strongly affected by, say, an abundance of poorly performing lenders in that particular MSA.

It is reasonable to begin with simple tests of association, like the chi-square, which are easy to calculate and can give a useful overall view of disparities in the market as a whole. As we have seen above, however, sample size requirements can be steep when dealing with rare events. If these tests give evidence that defaults are not distributed independently of tract or lender, it makes sense to turn to an examination of individual tracts and lenders, again starting with an analysis based on counts of loans and defaults. Simple tests based on the binomial, like those used above, are again easy to calculate and should reveal the particular tracts or lenders where problems may lie. It should be noted, however, that although starting with tests based on the statistical analysis of raw default rates minimizes data requirements, using such an initial screen could mask poor performance by lenders that operate in low-default tracts.

One should expect to find that some areas or lenders will fail the statistical tests; that is, for some areas and lenders there will be a low probability that as many defaults as are observed would occur by chance alone. One should not jump to the conclusion that such areas or lenders are necessarily problem prone. Rare events do occur and, especially when the number of tracts or lenders examined is quite large, one can expect to find that exceptional cases do indeed occur. Thus, what may appear unlikely to be chance does indeed occur by chance.

To help isolate tracts or lenders for which there really are problems, it may help to look for evidence of longer lasting effects. The assumption is that transitory problems will be viewed quite differently than more permanent difficulties, in part because temporary difficulties will be of less importance, and in part because they may be less amenable to remedial action. Hence, it may be informative to see if the same areas or lenders appear to have difficulties in more than one time period, restricting additional attention to those that appear to have problems more consistently.[56]

Those areas or lenders that appear to have longer lasting deficiencies may be subjected to additional analysis to see whether there are plausible reasons for the observed disparities in default activity that do not reflect poor underwriting or servicing. That is, the lack of controls for relevant factors makes raw differences alone rather irrelevant. Thus, one should examine intertract and interlender differences systematically by including other measurable default-related factors to see if such factors explain away observed differentials. Statistical procedures such as logit (which was used in this study), probit, or even linear regression may be applied to this problem at the level of the individual MSA.

There is no guarantee that estimation of statistical models for a single MSA will yield

[56] In the empirical work above, we looked for additional evidence on even those areas or lenders that failed the statistical tests for one origination year only. The reason we did so, however, is that we were concerned that intertract or interlender patterns in incomplete default data could cause some areas or lenders to appear acceptable when they are not.

appropriate estimates for key parameters, especially when the time period for originations is quite limited as well. The logit results for individual MSAs, presented in Appendix B, show many instances in which we obtain strange estimates for parameters for which the pooled logits, which utilize evidence across MSAs as well as within, obtain reasonable estimates. In such an event, one alternative is to use national samples to obtain parameter estimates that can be used as constraints in the MSA-level analysis.

Notice that when performing a more detailed statistical analysis, there is an advantage in deviating from what was done in the current study by making separate indicators for each individual tract or lender that has been singled out as potentially problematic. If complete separation is infeasible or undesirable (because, say, there are many tracts in the potentially high-default group), separation into small groups of tracts is an appealing alternative. Treating these areas or tracts as separate in the statistical procedures will make it possible to avoid having to make a blanket judgment on all potential problem areas together and instead permit one to single out those individual tracts or lenders that still appear troubled even after controlling for other factors.

Notice also that a more systematic study of default that controls for a variety of default-related factors could also serve to identify tracts or lenders that are under performing despite the fact that their raw default rates do not fail statistical tests. Picking these out would require a study of residuals by tract or lender.

6.3. Statistical Analysis of Lending or Servicing Practices

Whether or not the more detailed statistical analysis shows that interlender differentials vanish when additional factors are properly accounted for, one would want to see if lenders are following underwriting guidelines. Doing so, however, will require more data on underwriting factors (*e.g.*, on credit history) than were available in the current study or are generally available in FHA data files. The advent of automated underwriting systems may make such data more readily available for analysis. If the evidence is that underwriting guidelines are followed, and properly accounting for these and other factors explains away interlender and intertract differentials, then attention could turn to the possibility that underwriting criteria should be changed to include new factors or alter the tradeoffs permitted among existing underwriting factors.

Even if tightening underwriting criteria could be expected to reduce default rates, both overall and for specific areas and lenders, it does not of course follow that FHA should take such a step. FHA serves less affluent borrowers, thus extending home ownership to those who are less well served by the conventional market. Presumably, the benefits of homeownership in and of themselves serve to improve neighborhoods by promoting stability. Although reducing the risk in FHA lending by raising loan qualification standards can be expected to reduce default rates, it can also be expected to reduce FHA's ability to support the market that it has historically served. The result may be a lower default rate, but also reduced homeownership rates and thus reduced neighborhood stability. Thus, there is a tradeoff inherent in policy choices.

There is a very real possibility that unexplained differences among tracts and lenders will remain even after performing statistical analysis designed to adjust for a host of other factors. As in the current study, one may find that the impacts of tracts and lenders are reduced and in some cases vanish completely, but substantial estimated impacts remain for some lenders and some tracts. For a variety of reasons, this kind of result is not unexpected. Various underwriting factors may not be observed or they may be observed with considerable error, or other determinants of default may go unmeasured, and these factors may not be distributed evenly across tracts or lenders. In addition, analysts may not be able to reduce underwriting guidelines to simple formulas dealing only in observables. Underwriter judgement in particular may be difficult to quantify.

In the event that tract and lender impacts remain, additional avenues of inquiry could also be pursued. Data permitting, one could obtain measures of a variety of aspects of servicing performance --- intervals of delinquency prior to foreclosure, for example --- and employ these data to analyze differences in behavior across lenders. Pursuit of this idea in the current study, which raised serious questions of data quality, did not reveal any important and consistent differences in the speed with which high-default lenders intervened or the frequency with which the intervention took the form of an offer of an alternative to foreclosure.

Provided one had the resources, one could, of course, go further. One could, for example, sample application records from suspect lenders and use these to perform statistical analyses designed to see whether underwriting guidelines are followed. Such prospects are well beyond the scope of the current study but could be pursued elsewhere.

REFERENCES

Fleiss, J.L. *Statistical Methods for Rates and Proportions.* New York: John Wiley & Sons (1981).

Holloway, Thomas M., Gregor D. MacDonald, and John W. Straka,"Credit Scores, Early-Payment Mortgage Defaults, and Mortgage Loan Performance." Prepared for 1993 ORSA/TIMS Joint National Meeting. Phoenix, Arizona: November 1, 1993.

National Training and Information Center. "The Devil's in the Details: An Analysis of Federal Housing Administration Default Concentration and Lender Performance in 20 U.S. Cities." October, 1997.

Neal, Sue G. "Review of the Mortgage Default Literature," U.S. Department of Housing and Urban Development, Office of Policy Development and Research, unpublished working paper (December 1989).

Quercia, J.M. and Michael A. Stegman. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing and Research, 3, no. 2* (1992):341-80.